Immunology Research Tools and Resources

National Center for Biotechnology Information ■ National Library of Medicine ■ National Institutes of Health ■ Department of Health and Human Services

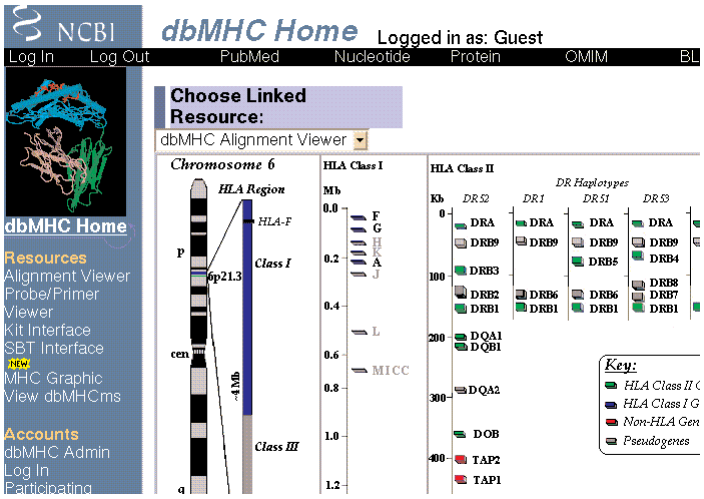
The National Center for Biotechnology Information (NCBI) supports and distributes a variety of resources for the medical and scientific communities. The following services provide data access and analytical tools to immunology and disease-related subjects.

dbMHC

NEW

The dbMHC database provides an open, publicly accessible platform for DNA, and clinical data related to the human Major Histocompatibilty Complex (MHC). The need to share research and clinical data focused on the MHC has lead to a series of meetings at the International HLA WorkShop & Congress (IHWC). The data generated within these workshops has proven to be valuable for the international HLA research community. In order to make these data easily available, the NCBI has committed itself to build dbMHC in close cooperation with the 13th IHWC. The goal is to display and maintain a permanent archive that contains the data generated within the 13th IHWC. In addition the dbMHC will provide tools for further submission and analysis of research data linked to the MHC. The URL will be:

www.ncbi.nlm.nih.gov/mhc



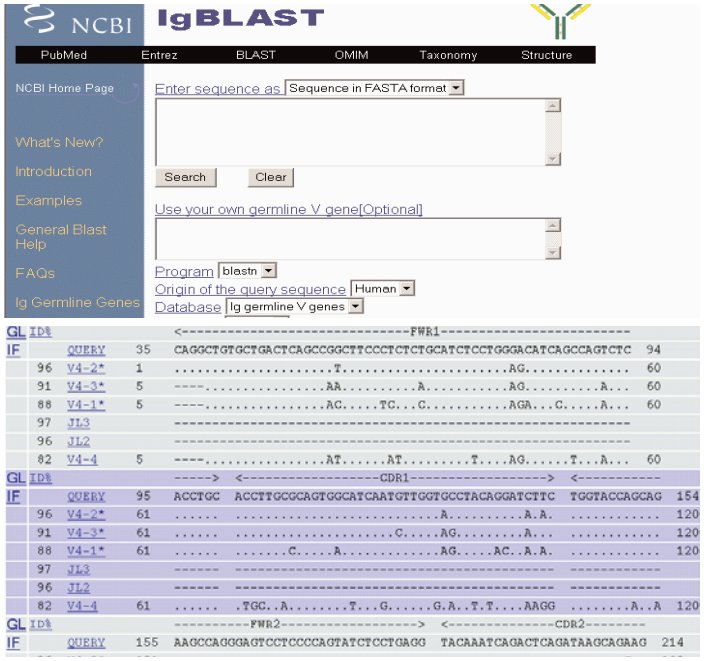
IgBLAST

There are many variants of BLAST (protein, nucleotide, iterative, domain etc.), including one specifically designed for immunology studies called IgBLAST. In addition to performing a regular similarity search, IgBLAST has three functions:

- Reports the three germline V genes, two D* and two J* genes that show the closest match to the query sequence.
- Annotates the immunoglobulin domains (FWR1 through FWR3) according to Kabat, et al .
- Matches the returned hits from the nr database to the closest germline V genes, thus making it easier to identify related sequences.

IgBLAST also annotates the query sequence, based on the domain information drawn from an alignment between the top-matched germline V gene and the query sequence, using the nomenclature of the Kabat Database of Sequences of Proteins of Immunological Interest. Such features as framework regions (FWR1, for instance) or complementarity-determining regions (CDR1, for instance) are delineated. The germline V-gene database currently contains Igh, Ig kappa, Ig lambda, and D and J genes from both human and mouse. The database can be viewed in the form of an annotated multiple sequence alignment by clicking on the appropriate database link in the sidebar of the IgBLAST page.

www.ncbi.nlm.nih.gov/igblast/



The top figure shows the IgBLAST query page. The bottom figure shows the IgBLAST output for a region of a rearranged lambda chain gene sequence taken from a patient with systemic lupus erythematosus (accession L28048). The predicated boundaries of FWR1 and CDR1 on the query sequence are indicated by IgBLAST on the lines marked "GL"

SAGE

One of the implementations heavily used in gene expression detection is Serial Analysis of Gene Expression (SAGE). It is a method of counting large numbers of mRNA transcripts by sequencing short tags, usually 10 base pairs (bp) in length. The analytical aspect involves mapping the tags to known genes and deciphering the degree of expression. SAGEtag to UniGene mapping (as well as vice versa- UniGene to SAGEtag mapping) examines data from SAGE libraries generated from the CGAP project and GEO libraries against the UniGene collection, including and in addition to genomic contigs, transcripts from GenBank, RefSeq and MGC, and dbEST. SAGE data has also been placed on genomic contigs from the human genome assembly, as the SAGE map on the Human Map Viewer. More details on SAGE and other useful products, such as xProfiler (which compares expression from two libraries), are located on the SAGEmap website at: www.ncbi.nlm.nih.gov/SAGE/index.cgi?cmd=expsetup

GEO

The Gene Expression Omnibus (GEO) project was developed in response to the growing demand for a public repository for high-throughput gene expression data. GEO provides a flexible and open design that facilitates submission, storage and retrieval of heterogeneous data sets from high-throughput gene expression and genomic hybridization experiments. The three central data entities of GEO are platforms, samples and series, and were designed with gene expression and genomic hybridization experiments in mind. A platform is a list of probes that define what set of molecules may be detected. A sample describes the set of molecules that are being probed and references a single platform used to generate its molecular abundance data. A series organizes samples into the meaningful data sets that make up an experiment. Scientists are encouraged to submit their expression data to GEO and detailed instructions are contained on the website. The GEO repository is accessible through our Web site at: www.ncbi.nlm.nih.gov/geo

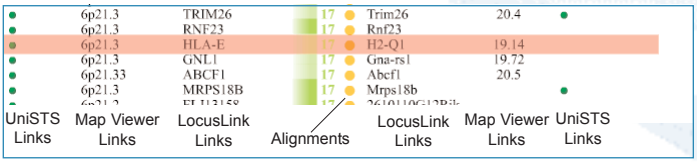
Map Viewer

The Human Genome Project (HGP) has produced physical and genetic maps in addition to the vast quantities of publicly available sequence data. To facilitate the analysis of complex genomic data, NCBI developed the graphical viewer called MapViewer.

The Map Viewer provides special browsing capabilities for a subset of organisms in Entrez Genomes. It allows you to view and search an organism's complete genome, display chromosome maps, and zoom into progressively greater levels of detail, down to the sequence data for a region of interest. The number and types of available maps vary by organism, and are described in the "data and search tips" file for each organism. If multiple maps are available for a chromosome, Map Viewer displays them aligned to each other based on shared marker and gene names, and, for the sequence maps, based on a common sequence coordinate system. One can see an example of a Map Viewer page to the right.

www.ncbi.nlm.nih.gov/mapview/

Map Viewer display showing SNPs, UniGene clusters and cytological mapping for HLA-E.



Human-Mouse Homology Map display for HLA-E.

OMIM

The Online Inheritance in Man (OMIM) is the database is of human genes and genetic disorders authored and edited by Dr. Victor A. McKusick and his colleagues at Johns Hopkins and elsewhere, and developed for the World Wide Web by the NCBI. The information in OMIM is textual in nature, and includes reports on genetic disorders and references. Data in OMIM can be queried via the Entrez search and retrievable system.

In addition to the ease of searching for phenotypes in OMIM, associated information from other NCBI resources are linked via a neighboring system to the current OMIM data. Check out OMIM in Entrez at:

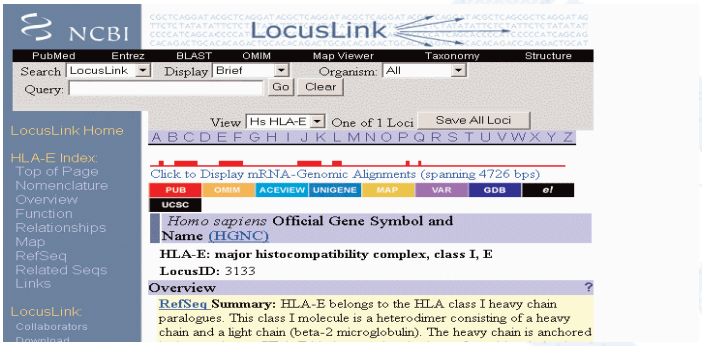
www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM

The results of the query "immune system disorders" are shown. Genomic loci are linked to the results if data are available.

LocusLink

LocusLink provides a single query interface to curated sequence and descriptive information about genetic loci. It presents information on official nomenclature, aliases, sequence accessions, phenotypes, EC numbers, MIM numbers, UniGene clusters, homology, map locations, and related web sites. LocusLink serves as a hub of information for loci from several model organisms: human, mouse, rat, fruit fly, and human immunodeficiency virus 1 (HIV-1).

www.ncbi.nlm.nih.gov/LocusLink/



LocusLink entry for the Human HLA-E gene. Nine links to additional sources of information or analysis tools are shown.

National Center for Biotechnology Information ■ 8600 Rockville Pike, Building 38A, Room 8N-803
Bethesda, MD 20894 Phone: 301-496-2475 ■ Fax: 301-480-9241 ■ email: info@ncbi.nlm.nih.gov
Web site: www.ncbi.nlm.nih.gov

revised 05/2003